The Power of Generative AI: A Review of Requirements, Models, Input-Output Formats, Evaluation Metrics, and Challenges

Abstract

Generative artificial intelligence (AI) has emerged as a powerful technology with numerous applications in various domains. There is a need to identify the requirements and evaluation metrics for generative AI models designed for specific tasks. The purpose of the research aims to investigate the fundamental aspects of generative AI systems, including their requirements, models, input–output formats, and evaluation metrics. The study addresses key research questions and presents comprehensive insights to guide researchers, developers, and practitioners in the field. Firstly, the requirements necessary for implementing generative AI systems are examined and categorized into three distinct categories: hardware, software, and user experience. Furthermore, the study explores the different types of generative AI models described in the literature by presenting a taxonomy based on architectural characteristics, such as variational autoencoders (VAEs), generative adversarial networks (GANs), diffusion models, transformers, language models, normalizing flow models, and hybrid models. A comprehensive classification of input and output formats used in generative AI systems is also provided. Moreover, the research proposes a classification system based on output types and discusses commonly used evaluation metrics in generative AI. The findings contribute to advancements in the field, enabling researchers, developers, and practitioners to effectively implement and evaluate generative AI models for various applications. The significance of the research lies in understanding that generative AI system requirements are crucial for effective planning, design, and optimal performance. A taxonomy of models aids in selecting suitable options and driving advancements. Classifying input–output formats enables leveraging diverse formats for customized systems, while evaluation metrics establish standardized methods to assess model quality and performance.